

# To the summit and beyond: Tracing the process and impact of collaborative performance summits

Scott Douglas<sup>1</sup> | Chris Ansell<sup>2</sup>

<sup>1</sup>Utrecht School of Governance, Utrecht University, Utrecht, The Netherlands

<sup>2</sup>Charles and Louise Travers Department of Political Science, UC Berkeley, Berkeley, California, USA

## Correspondence

Scott Douglas, Utrecht School of Governance, Utrecht University, Utrecht, The Netherlands.  
Email: [s.c.douglas@uu.nl](mailto:s.c.douglas@uu.nl)

## Funding information

Idea Generator, Grant/Award Number: NWA.1228.192.190

## Abstract

Interactive routines such as collaborative performance summits are thought to help collaborating organizations assess and improve their performance. However, there is little systematic evidence to substantiate this claim. This study leverages a longitudinal dataset to examine the summit process and identify the difference between summits that have an impact on performance and those that do not. The study explicates the assumed causal process and traces 18 partnerships as they prepare, conduct, and follow-up a summit. The analysis provides evidence for the positive impact of summits, but also shows that the process unfolds differently than expected. Neither the range of performance issues that actors bring to the summit nor the intentions for change they formulate at the end of the meeting are key differentiators. The hallmark of impactful summits emerges to be a large share of participants gaining comprehensive insights. These findings have implications for collaborative performance management research and practice.

## Evidence for practice

- Researchers argue that organizing collaborative performance summits could help the partners in a collaboration to better understand and improve their joint performance.
- This study of organizations collaborating on literacy programs explores the difference between summits that do or do not lead to performance improvement.
- The key characteristic of impactful summits is that a large share of participants report learning about the full breadth of the performance of the collaboration.
- To achieve this collective learning, summits should not be approached as purely technocratic information exchanges, but rather as interactive learning opportunities.
- The current study focuses specifically on the characteristics of the summit process itself, practitioners preparing summits will have to consider the characteristics of the wider collaboration to determine what type of summit will work best in their context.

## INTRODUCTION

Addressing the causes and consequences of a myriad of societal challenges—such as climate change, poverty, or pandemics—requires close collaboration between public, private, and community organizations (Head & Alford, 2015).

However, collaborating actors often struggle to get a grip on their collective performance, making it hard for them to assess their current progress and identify steps toward future improvement (Emerson & Nabatchi, 2015; Kroll, 2002).

Problems associated with performance management in the public sector in general—such as the difficulty of

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Public Administration Review* published by Wiley Periodicals LLC on behalf of American Society for Public Administration.

formulating and measuring performance indicators (Van Dooren et al., 2015)—are compounded by the specific challenges of collaborative governance settings—such as the emergent nature of goals and the divergent perspectives on success and failure among the various partners (Emerson & Nabatchi, 2015; Provan & Milward, 2001).

A key finding in the extensive research on performance management within organizations is that interactive dialogue routines, such as learning forums and performance reviews, can help actors to make sense of complex performance information and formulate potential improvement actions (James et al., 2020; Moynihan, 2005; Weick, 1995). Several authors have begun to explore the application of such interactive routines for improving the performance management of collaborating organizations (e.g., Bryson et al., 2020; Douglas & Ansell, 2021; Kroll, 2002; Laihonon & Mäntylä, 2017).

The current literature can provide a detailed account of the assumed causal process of interactive routines, but offers very little systematic, comparative, and longitudinal evidence to demonstrate that this process occurs as theorized and leads to performance improvement. On the whole, the empirical evidence is largely limited to single or small-N qualitative case studies and does not systematically track the performance impact of collaborative performance dialogues over time. This deficiency makes it hard to establish whether interactive routines make a difference.

This study systematically examines how interactive routines do or do not impact the performance of collaborations. The study specifically analyses the process and impact of a collaborative performance summit, which is an interactive dialogue routine where collaborative governance partners meet to exchange performance information, examine their progress, and explore potential improvements (Douglas & Ansell, 2021).

The study first synthesizes the current literature to outline the theoretical causal process of a summit, laying out the assumed sequential process from preparation and conduct of the summit to follow-up and subsequent performance impact. The study leverages a unique longitudinal dataset that traces 18 collaborations working to reduce adult illiteracy as they go through the summit process.

The findings provide evidence for the potential positive impact of summits on performance, but also show that the process unfolds differently than expected. The range of performance issues that actors bring to the summit or even the intentions for change they formulate at the end of the meeting are not key differentiators. The key difference between the summits that have a performance impact and those that do not is the share of participants that gain insights at the summit. Moreover, only those summits where participants gain insights on multiple dimensions of performance end up boosting improvement. The findings raise new

questions for collaborative performance summits in particular and collaborative performance management in general.

## THEORETICAL FRAMEWORK

### The complex nature of collaborative performance

A simple definition of performance would be the extent to which an organization accomplishes its goals (Gerrish, 2016). This definition requires some refinement in the case of collaborative governance, as the focus is not on a single organization but “a collective decision-making process based on more or less institutionalized interactions between two or more actors that aims to establish common ground for joint problem solving” (Douglas et al., 2020).

Collaborative governance emerges in the face of complex challenges where the ultimate goal (e.g., sustainable growth, healthy children, world peace) can rarely be fully accomplished (Head & Alford, 2015). Moreover, measuring goal fulfillment can prove practically unfeasible, as discovering the size and shape of the societal problem is part of the collaborative challenge (Ansell & Gash, 2008). And above all, different actors in the collaboration will have different definitions of performance and these goals may shift over time (Vangen & Huxham, 2013).

For example, the collaborations in this study try to reduce functional illiteracy in their communities. Across OECD countries, about one in five adults struggle to read messages from their government, employer, children’s teacher, or healthcare provider (OECD, 2019). This problem affects different groups in society (e.g., immigrants raised in a different language, low-income households with poor access to resources, people with learning difficulties), with every group requiring a customized solution. Total eradication of illiteracy is the ultimate goal, but tracking this goal only would give partners only a very broad and slow-moving performance indicator. Moreover, people actively hide their illiteracy out of shame, making it difficult to measure the size of the problem. Recognized illiteracy rates may even go up as the partners get better at spotting it.

Above all, the various partners involved in literacy programs will often have different definitions of good performance. Community colleges might measure success by the amount of people that pass their formal literacy exams, while libraries want people to engage with reading in whatever way suits them. Employers want their workers to be able to understand technical manuals, while illiterate people themselves may simply want to be able to read bedtime stories to their kids. Moreover, these goals will shift as new problems are identified (e.g., the link between poor health and illiteracy) and new insights are gained (e.g., new training methods).

These complexities do not make performance a useless concept for collaborations, but do mean that collaborative performance needs to be considered in all its nuances: (Choi & Moynihan, 2019). Authors such as Provan and Milward (2001), Emerson and Nabatchi (2015), and Douglas and Ansell (2021) distinguish between different dimensions of collaborative performance, such as effectiveness, adaptability, and legitimacy. In line with these frameworks, this study distinguishes between the *operational* dimension of collaborative performance (e.g., the quality of the service the partners are jointly delivering to illiterate people), the *strategic* dimension, (e.g., the extent to which partners can formulate goals), and the *constitutional* dimension of performance (e.g., the extent to which partners can successfully structure their coordination). Together, these dimensions provide insight into the ability of a collaboration to deliver on its ambition.

### The need for interactive routines such as collaborative performance summits

The innate complexity of performance in the public sector initially led to a push for more and more performance indicators (Van Dooren et al., 2015). However, research on performance management within organizations established that performance information is rarely used if actors are not also provided with interactive routines to jointly make sense of all this data and explore what steps to take next (Moynihan, 2005; Moynihan & Kroll, 2016; Weick, 1995). Effective performance management is more about crafting dynamic systems (Bianchi, 2016) and creating usable knowledge (Lindblom & Cohen, 1979), than about designing static measurement systems and producing summary judgments.

For collaborative governance, the push for interactive performance routines aligns with the realization that collaborations need processes for dealing with the cognitive, strategic, and institutional complexities they face (Koppenjan & Klijn, 2004). And that the processes for formulating goals and defining next steps are constantly ongoing and evolving within collaborative governance arrangements (Ansell & Gash, 2008; Vangen & Huxham, 2013).

This study focusses specifically on interactive routines in the shape of collaborative performance summits (Douglas & Ansell, 2021). During such summits, the partners in a collaboration come together to jointly discuss their collective progress and what they should do next. The concept of a summit is related to the concepts of a learning forum (Moynihan, 2005), performance dialogue (Laihonen & Mäntylä, 2017), PerformanceStat meeting (Behn, 2014), or a forum (Bryson et al., 2015).

The literature identifies various ways in which interactive routines such as summits could have an impact on collaborative performance. Summits can serve as planning meetings that help actors to consider alternative directions and break down big tasks into manageable

steps (e.g., George et al., 2019). Summits can act as political arenas where partners can build consensus for action (e.g., Bryson & Crosby, 1993). Or summits can serve as learning opportunities by offering actors a safe place for making sense of complex performance information and figuring out what to do next (e.g., Gerlak & Heikkila, 2011).

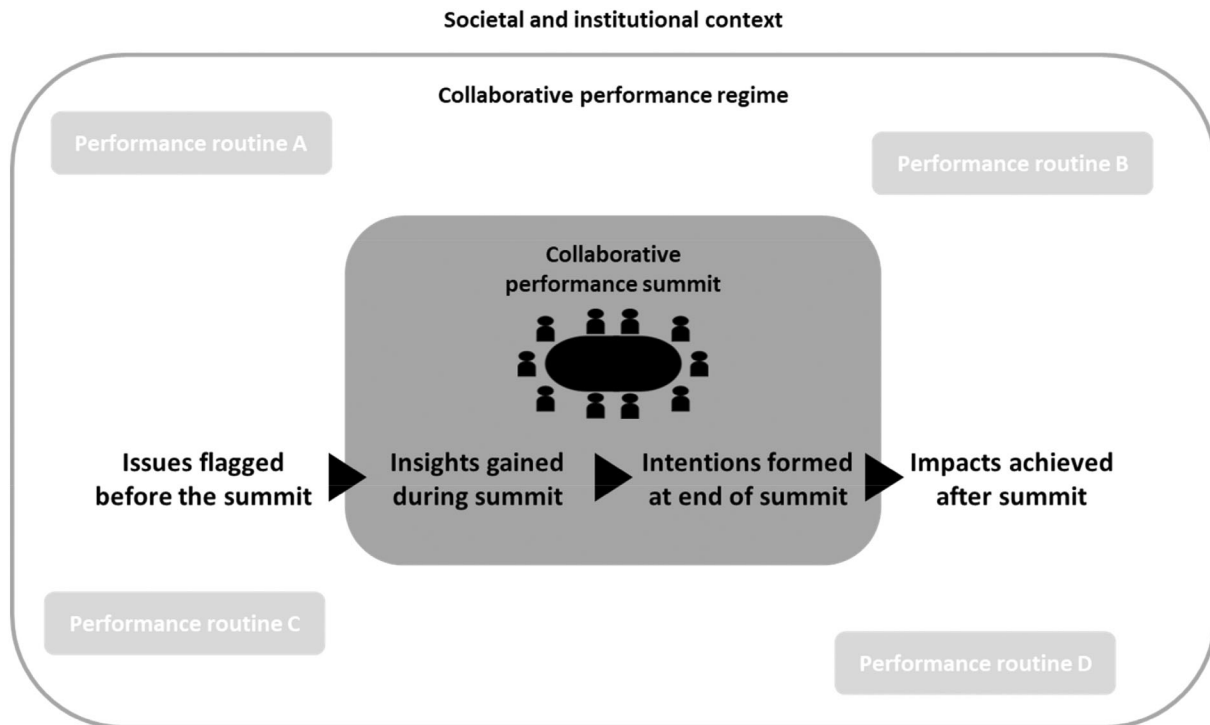
### The assumed causal process of collaborative performance summits

Collaborative performance summits as described in the literature are here understood as intentional causal processes (Collier, 2011). The core premise is that the sequential steps undertaken by actors in the preparation, conduct, and follow-up of a summit (the chain of causal links) will lead to an improvement in collaborative performance (the ultimate effect) (George & Bennett, 2005).

Authors may differ in the precise steps they think this causal process consists of and the tightness of the link between different steps. Behn (2014), for example, approaches the organization of a performance summit as a military campaign where a regimented sequence of actions should lead participants to logical conclusions about the data they are presented with. Moynihan (2006) also thinks that learning forums require careful preparation, but suggests there will always be an element of randomness in the connection participants make between information and subsequent action.

In line with Gerlak and Heikkila's (2011) description of collective learning as a process of stock-taking, sense-making, action-planning, and improvement, the causal process of a summit meeting is here broken down into four consecutive steps: first, participants bring a range of *performance issues* to the summit, identifying bottlenecks or opportunities they see in the current functioning of the collaboration; second, participants discuss this information during the summit and gain new *performance insights* into how the partnership is doing; third, these insights are translated into *intentions* for change to improve the current performance; finally, after the summit, these changes should lead to a positive *performance impact*.

Although the causal process described here posits a link between these four sequential steps, this relationship is not assumed to be deterministic. Like most social phenomena, the process of summits is influenced by a myriad of factors (Yin, 2009). For example, the dynamics of the summit will be shaped and constrained by the overall performance regime of the collaboration (Douglas & Ansell, 2021). This context influences which actors are likely to be involved in the summit, what other routines will play out alongside the summit (e.g., bilateral meetings), and what actions are deemed appropriate. Figure 1 presents the four steps of the assumed causal process placed within the wider collaborative performance regime.



**FIGURE 1** Overview of the causal process of summits and the wider performance regime

## Tracing the causal process and impact of collaborative performance summits

Having explicated how the causal process of a summit will unfold in theory, the question arises whether this process does indeed occur like this in practice. A few studies empirically describe the preparation and conduct of summits (e.g., Bryson et al., 2015; Douglas & Schiffelers, 2021; Laiho-nen & Mäntylä, 2017). However, the available studies are either theory-generating rather than theory-testing exercises, limited to single or small-N case studies, or crucially do not systematically examine the impact on performance of the summit beyond the immediate aftermath of the meeting. This scarcity of evidence makes it hard to demonstrate how summits make a difference.

This study leverages a unique dataset following 18 partnerships as they prepare, conduct, and follow-up a collaborative performance summit. This data allows for the examination of multiple questions. First, the analysis can explore whether the sequential links between the various steps in the summit process unfold as theorized. Second, the analysis can compare the pathways of different cases, examining what makes the difference between impactful and non-impactful summits. Third, the analysis can probe the role the different operational, strategic, or constitutional dimensions of performance play in this process. On the whole, the results of this study can support or challenge the faith researchers have in the beneficial effects of interactive routines on the performance of collaborations.

## METHODOLOGY

### Local collaborations to reduce adult illiteracy

The case material for this study is drawn from 18 local partnerships in the Netherlands that seek to reduce adult illiteracy. Across the OECD, about 5% of adults score below the lowest level of text comprehension, while a further 15% struggle with basic reading and writing tasks. This functional illiteracy severely limits their ability to correspond with the government, control their finances, make healthy choices, navigate the digital world, or generally participate in society (OECD, 2019). Many of the people affected by illiteracy are ashamed to admit their problem and there can be many causes (previously undiagnosed dyslexia, early school dropout, having being raised in a different language, etc.). This necessitates a close collaboration between different frontline services to identify people in need (e.g., family doctors, social workers, volunteers) and offer customized support (e.g., a community college to offer a formal course, a local charity group to provide a reading buddy).

The national government has tasked municipalities with coordinating and financing the local literacy program. Although local governments in the Netherlands have a relatively large administrative capacity, the complex nature of the problem still means they have to collaborate with a wide array of public, private, and community organizations. Over the previous years, the local governments were able to sign up multiple partners

**TABLE 1** Characteristics of the 18 summits

#	Population	Summit participants							
		Local govt.	Comm. College	Library	National foundation	Charity	Job agency	Social services	Other
A	20,000	●	●	●	●	●	●	● x3	● x4
B	30,000	●	●	●	●			●	
C	35,000	●	●	●	●		●	● x3	
D	20,000	● x2	●	●	●			●	
E	120,000	●	●	●	●		●	● x3	
F	650,000	● x19	●	● x5	●				
G	90,000	●	●	●	●		●	● x4	
H	10,000	●	●	●	●			●	
I	30,000	●	●	●	●			● x4	
J	150,000	● x3	●	● x2	●			● x2	
K	55,000	●	●	●	●			●	● x6
L	30,000	●	●	●	●			●	● x6
M	40,000	●	● x2	●	●			●	●
N	40,000	●	●	●	●			●	
O	25,000	●	●	●	●			●	● x4
P	5,000	●	●	●	●			● x2	● x4
Q	20,000	● x3	●	●	●		●	● x3	● x4
R	125,000	●	●	●	●			● x2	●

to 'literacy pacts'. These pacts were usually framed in very general language, although in some localities formal collaborative performance regimes were established with specific goals. Across all municipalities, however, partnerships had to deal with the absence of reliable measures of the illiteracy rate, disagreement between partners about what should be prioritized, and uncertainty about what approach to illiteracy worked best.

### Following 18 collaborative performance summit processes

The researchers were previously asked by the Foundation for Literacy, a charity financed by the national government, to develop a summit format that would facilitate a dialogue between all the partners involved in a local literacy program. After an initial pilot phase (see Douglas & Schiffelers, 2021 for an analysis of these pilot cases), this format was made available to all 300+ municipalities in the country, with the national foundation providing a moderator to lead a summit meeting and the researchers remaining involved at arms-length to collect data about the preparation, conduct, and impact of the summits. Over the course of 18 months, 25 municipalities volunteered to organize a summit. These municipalities were all located in urban/semi-urban settings and varied from small to medium-sized towns (in one case, the summit brought together multiple municipalities working within the same region).

In each municipality, the local civil servant responsible for the literacy drive and a regional expert of the

national literacy foundation determined which partners should be invited to the local summit. They selected those partners that could provide an insight into the impact of the literacy program and would be essential in improving the program. Representatives of the municipality, library, community college, and a local representative from the national foundation participated in each summit. The social services and charities (e.g., the Salvation Army, local volunteers) were represented in most local summits. Job agencies attended half the summits and local employers four of them. In some cases, participating organizations choose to come with multiple representatives (see Table 1).

Every summit followed the same format: The participants were first asked by the moderator to pick a picture from a range of photos that best represented the current state of their collaboration (e.g., a decrepit house, a clockwork mechanism, a group of monkeys). Each participant was then given a minute to explain why they choose that particular picture. The participants were then presented with the anonymized results of a survey of the participants conducted prior to the summit, which detailed their aggregate assessment of crucial aspects of the collaboration (covering such elements as the clarity of goals, quality of language courses, and deployment of resources). Participant were asked to discuss this survey to better understand how the collaboration was doing. Each meeting concluded with an exploration of possible performance improvement actions. All summits lasted between one and a half and two and a half hours.



**TABLE 2** Coding for performance issues, insights, intentions, and impacts

<b>Issues flagged by collaboration partners in pre-summit survey</b>	<b>Insights reported by collaboration partners in end-of-summit survey</b>	<b>Intentions formed by collaboration partners in end-of-summit survey</b>	<b>Impacts observed by illiteracy expert 6 to 12 months after summit</b>
Partners were asked to agree or disagree with the below statements on a 5-point scale. If they Disagreed or Strongly Disagreed, these were marked as flagging a performance issue <sup>a</sup>	Partners were asked to agree or disagree with each statement below on a 5-point scale. If they Agreed or Strongly Agreed, these were marked as having gained an insight	Partners were asked to agree or disagree with each statement below on a 5-point scale. If they Agreed or Strongly Agreed, these were marked as projecting intentions for change	The regional literacy expert was asked to agree or disagree with the below statements on a 5-point scale. If they Agreed or Strongly Agreed, they were marked as observing impact
<b>Operational issues</b>	<b>Operational insights</b>	<b>Operational intentions</b>	<b>Operational impacts</b>
<ul style="list-style-type: none"> <li>The recruitment, guidance, support, and examination of clients is going well</li> </ul>	<ul style="list-style-type: none"> <li>I have received new and relevant information during the summit</li> </ul>	<ul style="list-style-type: none"> <li>The summit will lead to practical improvements</li> </ul>	<ul style="list-style-type: none"> <li>The summit has led to improvements in the delivery of the shared program</li> </ul>
<b>Strategic issues</b>	<b>Strategic insights</b>	<b>Strategic intentions</b>	<b>Strategic impacts</b>
<ul style="list-style-type: none"> <li>All partners have a clear view of the ambitions of the local collaboration and the goals we want to achieve</li> </ul>	<ul style="list-style-type: none"> <li>I gained a clearer understanding of the goals of the local collaboration during the summit</li> </ul>	<ul style="list-style-type: none"> <li>The summit will lead to changes in the goals and ambitions of the collaboration</li> </ul>	<ul style="list-style-type: none"> <li>The summit has resulted in the adjustment of the goals and ambitions of the collaboration</li> </ul>
<b>Constitutional issues</b>	<b>Constitutional insights</b>	<b>Constitutional intentions</b>	<b>Constitutional impacts</b>
<ul style="list-style-type: none"> <li>The coordination is effective in convening, stewarding, mediating and catalyzing between partners</li> </ul>	<ul style="list-style-type: none"> <li>I gained a better understanding on the quality of our collaboration during of the summit</li> </ul>	<ul style="list-style-type: none"> <li>The summit will lead to a new relationship dynamic between the partners</li> </ul>	<ul style="list-style-type: none"> <li>The summit has resulted in improved coordination of the collaboration</li> </ul>

<sup>a</sup>"I don't know" was also an option and coded as a 0, indicating a need for learning.

## Tracing the causal process of the summit

In line with the causal process described above, data was collected on the performance *issues* that actors flagged before the summit, the *insights* they gained at the summit, the *intentions* for change that each participant took away from the summit, and the performance *impact* observed after the summit. The data also tracked whether these views pertained to the operational, strategic, or constitutional dimensions of performance.

Concerns about performance (*issues*) were identified by directly asking the partners about them in an e-survey 2 weeks before the summit. This survey was developed by the researchers in collaboration with the Literacy Foundation and piloted in a separate set of eight municipalities to make sure the questions were clear. This pre-summit survey also provided the input for the summit itself as the partners discussed the anonymized results during the meeting. An e-survey conducted right at the end of the summit solicited feedback from participants about what they had learned (*insights*) and what changes they expected over the coming months (*intentions*). Performance results (*impacts*) were not captured by asking the local partners themselves since this might have returned biased results. Instead, the literacy expert of the national foundation working in the specific region in which the summit took place filled out an online survey rating the impact of the summit 6 to 12 months after the event.

Table 2 provides an overview of the exact questions that were put to the summit participants and regional literacy experts.

## Collecting data

The plan was to collect data for all 25 municipalities that organized summits in the 18-month research period. However, in three cases, the group ran out of time at the end of the summit and there was no end-survey. In two cases, only two of the summit participants completed the end-of-summit survey, providing too little data. In two final cases, the post-summit survey was missing as there was no literacy expert working in the region at the time. With the dataset for these seven cases incomplete, they were excluded from the analysis, yielding a total of 18 cases.

Table 3 provides a basic overview of the results, presenting the scores on performance issues, insights, intentions, and impact for each case averaged out across the operational, strategic and constitutional dimensions. The scores for *issues* reflect the percentage of participants in the collaboration that flagged serious or very serious performance concerns. The scores for *insights* reflect the share of summit participants that reported gaining insight at the summit. The scores for *intentions* reflect the share of summit participant that anticipated changes after the

**TABLE 3** Raw data on the issues, insights, intentions, and impacts observed per case

Case	Share of actors flagging issues in pre-summit survey	Share of participants gaining insights in end-of-summit survey	Share of participants projecting intentions in end-of-summit survey	Degree of impact observed by illiteracy expert 6 to 12 months after summit
A	65%	47%	33%	58%
B	48%	75%	67%	92%
C	61%	56%	67%	58%
D	28%	61%	56%	83%
E	34%	63%	50%	58%
F	36%	10%	19%	17%
G	28%	67%	69%	58%
H	54%	75%	75%	67%
I	63%	83%	58%	50%
J	28%	72%	83%	75%
K	52%	59%	67%	50%
L	63%	60%	53%	75%
M	64%	33%	56%	50%
N	24%	67%	67%	42%
O	53%	73%	67%	75%
P	49%	75%	92%	67%
Q	16%	37%	11%	42%
R	26%	30%	41%	25%

summit. The scores for *impact* reflect the assessment of the illiteracy expert of the performance impacts of the summit.

For example, in case A, 14 different participants filled out the pre-summit survey. Of these respondents, 60% thought the recruitment and support of illiterate people within the various programs was not going well (operational issues), 79% thought the shared goals were unclear (strategic issues), and 55% thought the coordination of the collaboration was ineffective (constitutional issues). Averaged out across the three performance dimensions, this made for an average of 65% of the respondents reporting issues.

At the end of the summit, 80% of the participants felt they had gained insights on the operational performance of the collaboration, 20% had gained insights into the strategic goals of the collaborations, and 40% on the overall dynamics of the partnership, making for an average of 47% participants reporting insights. A lower proportion (33%) projected *intentions* for the summit to lead to actual operational, strategic, or constitutional changes.

Six months later, the literacy expert active in the region concluded that there were no discernible *impact* on the operational performance (rating it a 3 out of 5, converted to a 50% score), some impact on the strategic dimensions of performance (giving a 75% for the realignment of the goals), and no discernible impact on the constitution of the collaboration (50%). The average of these three assessments was 58%. (Appendix A contains the

detailed scores for each case, listing the operational, strategic, and constitutional dimensions of performance separately.)

### Analyzing the data to probe the assumed causal process

The central ambition of this study is to examine whether the causal process of the summit as assumed in the theory unfolds as such in practice. The focus is on “the causal chain and causal mechanism – between an independent variable (or variables) and the outcome of the dependent variable” (George & Bennett, 2005, pp. 206–207). With a medium-N set of cases available, probing such causal inferences can be done through different analytical techniques, where each methodological option has its particular strengths and weaknesses.

One option would be conducting a Qualitative Comparative Analysis (QCA) of the available causal process observations to examine what configurations of factors need to be absent or present for the summit to have a positive impact on performance. Such an analysis might establish what combinations of conditions are necessary and/or sufficient for making a summit a success (Schneider & Wagemann, 2010).

However, importantly, the theory examined here presents the assumed causal process not only as *configuration of conditions* but as a *sequence of steps*. The identification of issues, the generation of insights, the

formulation of intentions, and the realization of impact are assumed to follow each other chronologically, with each stage shaping the next. QCA can be adapted to gain some insight in mechanisms over time, but the method is primarily designed for assessing configurations rather than sequences. QCA might be of better use when zooming out to probe the link between the summit process and its wider performance regime, for example, examining what role contextual conditions play in making a summit a success.

To properly examine the assumed sequence of steps *within* the summit process, this study opts for an analytical approach more closely related to process-tracing, focusing “on the unfolding of events or situations over time” (Collier, 2011). The analysis traces the 18 cases as they move from performance issues to insights, from performance insights to intentions, and from performance intentions to impact. In each step, the empirical evidence for the assumed causal link between the parts is examined. This approach allows the analysis to open up the black box of the summit process and zoom in on the assumed sequential nature of the process.

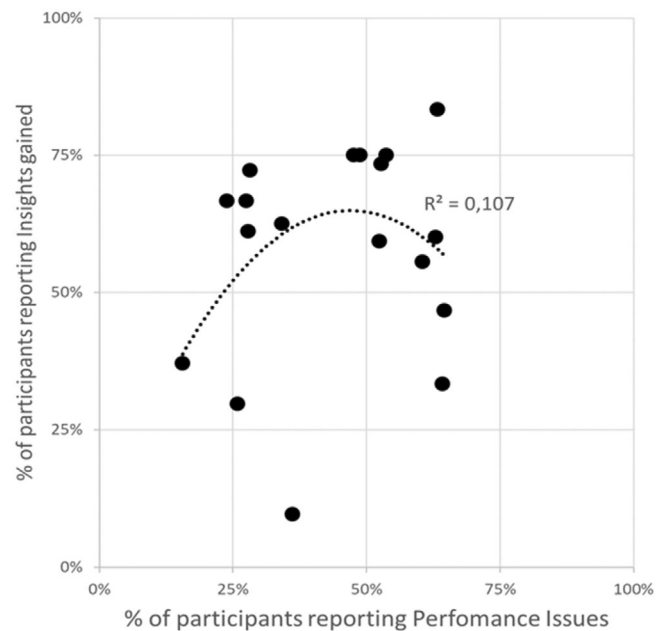
Alongside descriptive texts, simple scatterplot diagrams and  $R^2$  calculations will be used to explore how the different steps of the summit process shape each other and where cases deviate from the pathway as theorized. These visualizations and calculations are only meant to complement the more in-depth discussions of the scores of specific cases; they do not amount to full statistical evidence because the  $N$  is too small. Considerable attention is also paid to dissecting the case scores (either for specific actors such as the municipality or for the difference between operational, strategic, or constitutional dimensions of performance) to further probe the more nuanced dynamics at work. The Discussion will reflect on the upsides and downsides of the methodological choices made.

## RESULTS

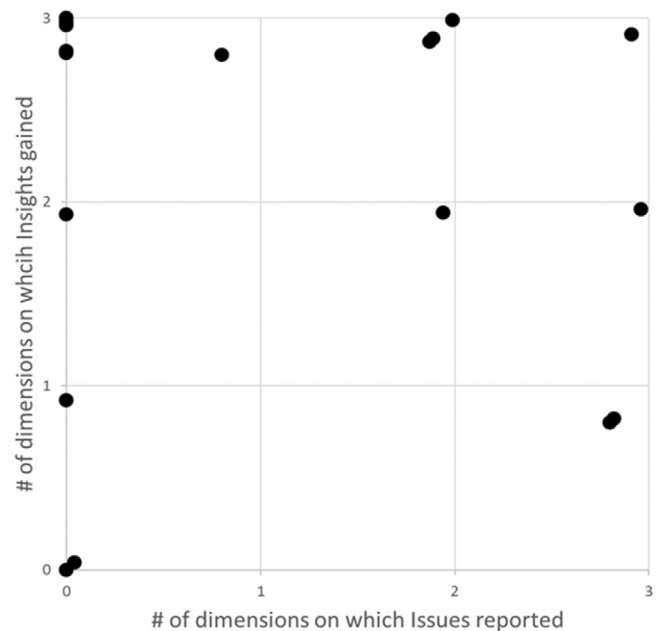
### From performance issues to insights: A scattershot translation

The participants were asked before the summit to rate the various aspects of their collaborative performance. This assessment varied greatly across cases. For example, only 16% of the partners in case Q felt there were significant performance issues, while 65% of the partners in case A felt the collaboration faced performance challenges.

At the end of the summit, participants were asked what performance insights they gained. Again, there was a large variation in the responses, although most partners reported gaining more insights than expected. For example, only 28% of the participants in case J saw performance issues before the summit, but 72% reported gaining relevant new insights after the summit.



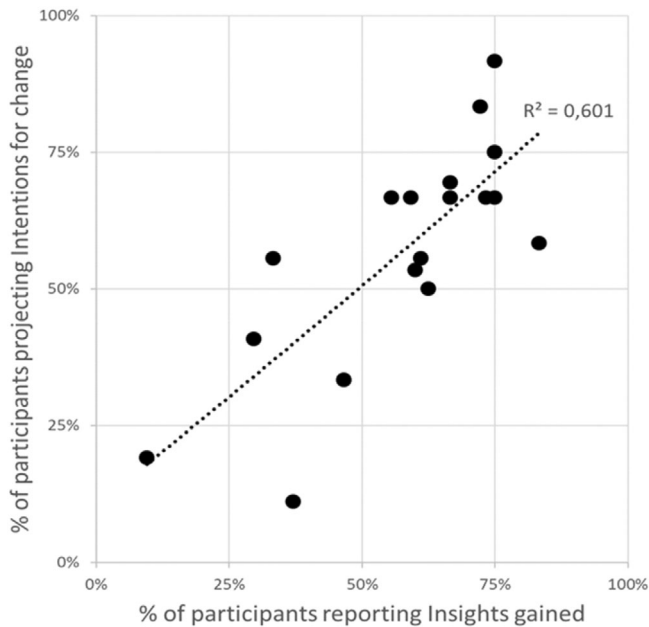
**FIGURE 2** Weak link between share of participants reporting performance issues and share of participants gaining insights



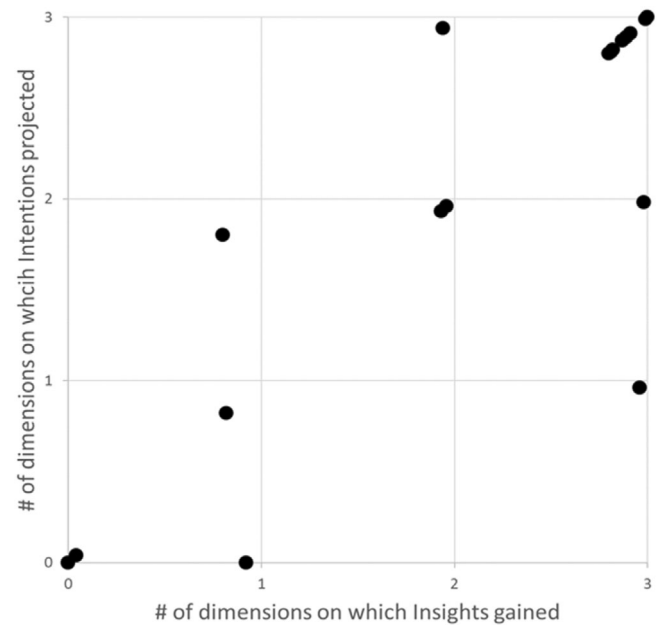
**FIGURE 3** Participants generally gain insights on more performance dimensions than they initially reported issues on

This pattern of ‘learning more than expected’ is repeated when examining the number of performance dimensions participants reported learning about. For example, in case D, partners had reported no significant issues on either the operational, strategic, or constitutional dimensions of performance. However, at the end of the summit, a majority of the participants in this case





**FIGURE 4** Close link between share of participants gaining insights and participants seeing intentions for change



**FIGURE 5** Close link between the number of dimensions on which insights are gained and number on which intentions are seen

reported having gained new insights on all three performance dimensions.

Figures 2 and 3 visualize the relationship between the issues reported and the insights gained. On the whole, a larger share of partners gained insights at the summit, and on more dimensions of performance, than they reported performance issues before the summit. There is still a small positive correlation between the share of participants flagging issues before the summit and the number of participants gaining insights at the summit. However, this effect is rather small and scattershot ( $R^2 = 0,107$ ). There are multiple cases where a low number of issues are identified but where summits end up producing a lot of insights, or vice versa. This effect seems to be less pronounced and even negated when a lot of issues are reported.

### From performance insights to intentions: A close translation

Next to asking participants what insights they gained, participants were also asked at the end of the summit what intentions for change they saw to improve the performance. Again, the responses varied. In case Q, only 11% of participants saw real intentions for change, with no majority of actors expecting change on either the operational, strategic, or constitutional elements of the collaboration. In case P, participants were nearly unanimous (92%) in seeing real intentions for change and thought this would affect all the performance dimensions.

The visualizations in Figures 4 and 5 suggest a strong link between the share of participants that report gaining new insights and the share of participants seeing intentions for change. As the share of participants gaining new

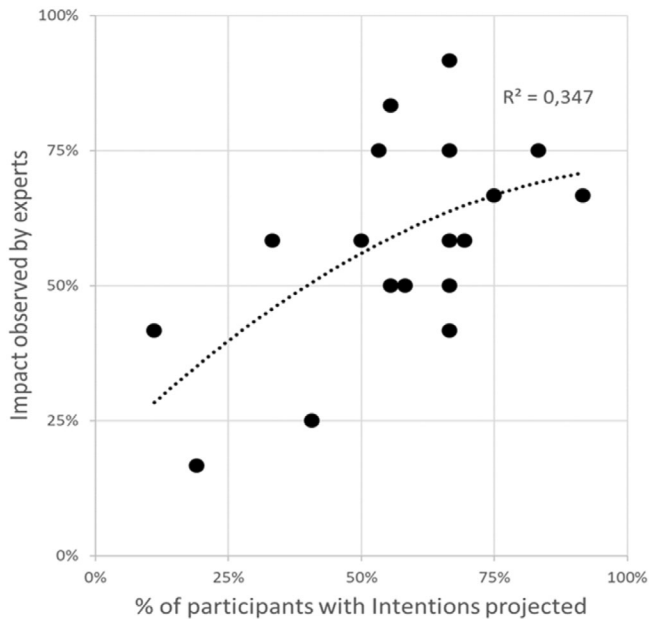
issues goes up, the share of participants projecting change tends to go up as well ( $R^2 = 0,601$ ). Similarly, if participants gain insights on more dimensions of performance, they are also likely to see intentions for change on more dimensions of performance. (Note that the participants were asked these two questions at the same time in the same survey, which may have bolstered the connection).

Interestingly, insights into a specific performance dimensions do not necessarily translate into intentions for change for this same performance dimension. For example, insights into strategic issues generally translate into intentions for making changes to the strategic priorities. However, this specific link between strategic insights and strategic intentions is weaker ( $R^2 = 0,340$ ) than the relationship between insights and intentions overall ( $R^2 = 0,601$ ). The relationship between operational insights and operational intentions, or constitutional insights and constitutional intentions is weaker still.

In multiple cases, insights on one dimension of performance actually led to intentions on different dimensions of performance. For example, in case A, the partners gained insights into the operational performance of the collaboration, and actually concluded that this could be best addressed by changing the constitution of the collaboration. This suggests that the link between insights and intentions is more complex.

### From performance intentions to impact: Lost in translation?

Six to twelve months after the summit, the regional literacy expert of the national foundation was asked to rate



**FIGURE 6** Positive link between number of issues on which participants made intentions and the expert observed positive impact

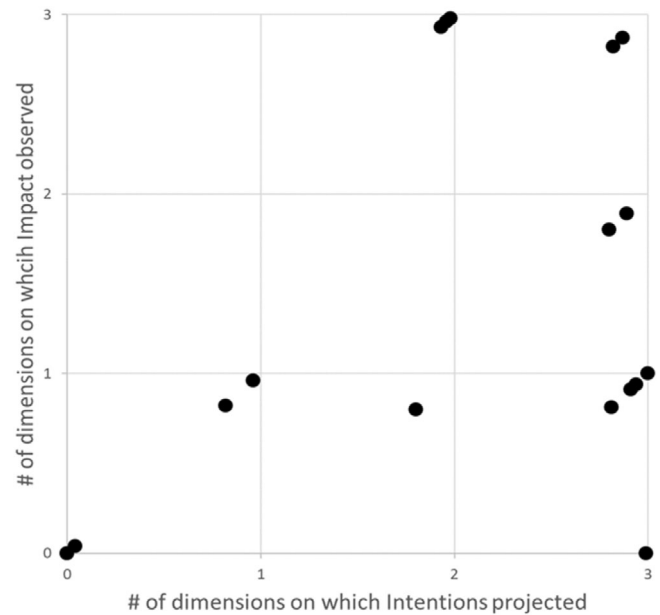
the impact of the summit on performance. On the whole, the share of participants that saw intentions for change at the end of the summit is positively correlated with the assessment of the impact by the regional expert after the summit (Figure 6).

For example, in case J, 83% of the participants of the summit saw intentions for change and the expert later concluded that there was improvement across the operational, strategic, and constitutional performance dimensions. On the whole, this relationship between intentions and impact is relatively close ( $R^2 = .347$ ), but cannot explain all the variety in the outcomes. For example, in case P, 92% of the participants saw intentions for change, while the experts rated the impact only at 67%. The reverse was true for case B, where only 67% of participants saw intentions for change, but the expert rated the impact at 92%. This suggests that other factors may help or hinder the eventual impact of the summits.

Figure 7 contrasts the number of dimensions on which participants expected change to the number of dimensions on which the expert observed impact. Interestingly, in almost all cases participants overestimate the number of dimensions in their intentions for change, realizing impact on fewer dimensions of performance than expected. This could point to a loss in the translation from intentions to implementation. A more positive reading would be that having ambitions across multiple dimensions is necessary for affecting change in at least some dimensions.

### From beginning to end: The difference gaining insights makes

The analysis has so far examined each of the separate links in the sequence of the summit process. It is also



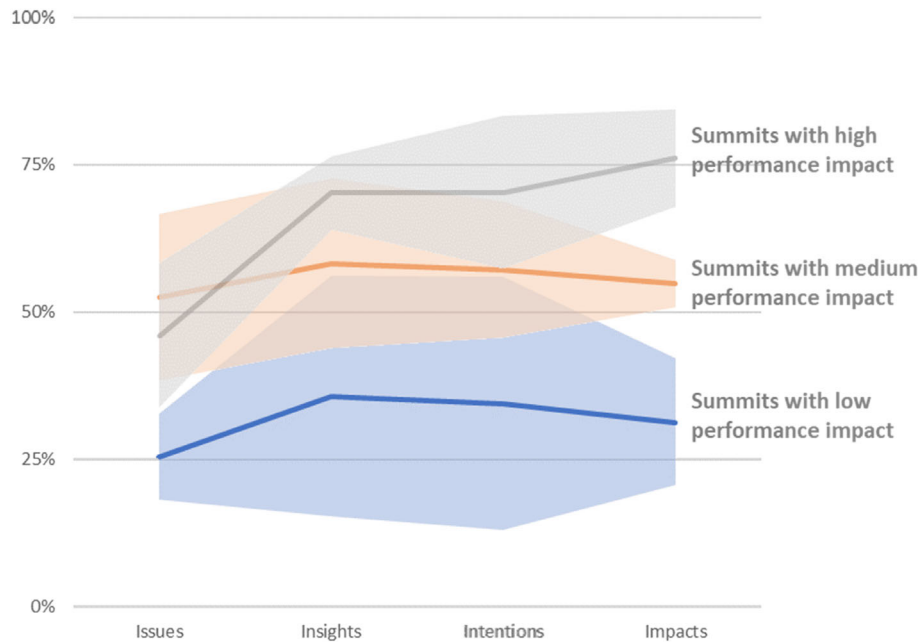
**FIGURE 7** Number of dimensions on which impact is realized is only high when intentions were made on multiple dimensions

important to consider the full pathway of a case as it moves through the summit process in its entirety. The pathways traveled by cases that end up realizing a large impact on performance can so be contrasted with the pathway of cases realizing only a moderate or low impact on performance.

Figure 8 visualizes the pathways of three different types of cases. The top range in gray reflects the pathway of the seven cases that end up with a high impact score, with the line charting their average score on each of the four parts of the process and the transparent band marking how much individual cases deviated from this average. The middle range in orange charts the course of the seven cases achieving a moderate impact, with the band marking the deviations from this average. The bottom range in blue charts the course of the four cases achieving a low impact score, with the transparent band highlighting the large deviations from the average score of this group of cases.

Interestingly, both the high- and moderate-impact cases generally start with around half of the partners flagging performance issues before the summit. The difference between the moderate versus high cases is the share of participants that report gaining new insights, with the high-impact group achieving a much higher share of participants learning than the moderate impact group. The translation from insights to intentions is then stable for both groups, with the high impact set attaining a higher score. The high-impact group makes some final gains in the translation from intentions to impact after the summit, while the moderate impact group loses steam in this step.

The four cases in the low impact group all start with only a small share of partners seeing performance issues



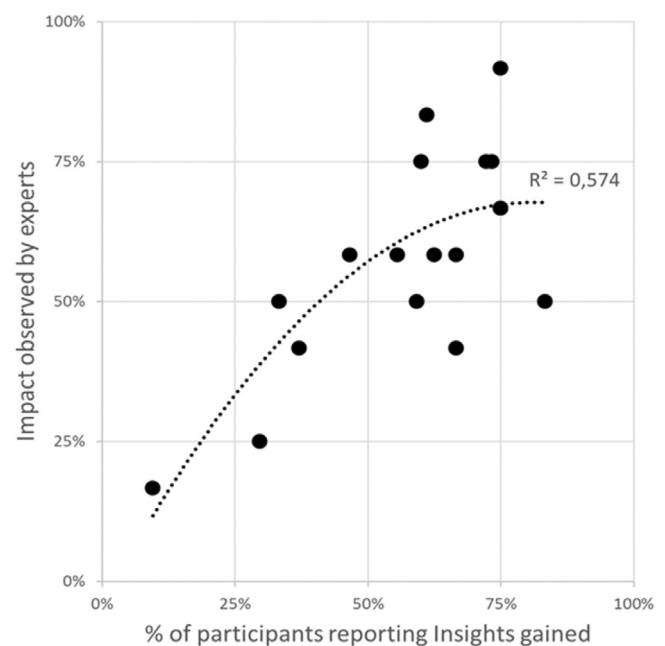
**FIGURE 8** Different pathways of high-, middle-, and low-impact summits

and then report the lowest share of participants gaining new insights at the summit. The subsequent pathways is more varied in this set of low impact cases. In case R, for example, the scores remain low across the whole summit process (around the 20% at each step). For case N, only 24% of the partners see initial issues, but 73% of the participants report gaining new insights and 67% seeing intentions for change. In the end, however, the case only achieves a 42% impact rating as the initial momentum for improvement seem to run into difficulties after the summit.

The difference that the share of participants gaining new insights makes for the impact of a summit is further emphasized by cross-checking the links between all steps of the summit process. This analysis shows a scattershot link between issues reported and intentions projected, and between issues reported and impact observed. However, the link between the share of participants reporting insights and the impact observed by the expert is very close (Figure 9). And again, the number of dimensions on which impact realized is only high in cases where insights were gained on multiple dimensions (Figure 10).

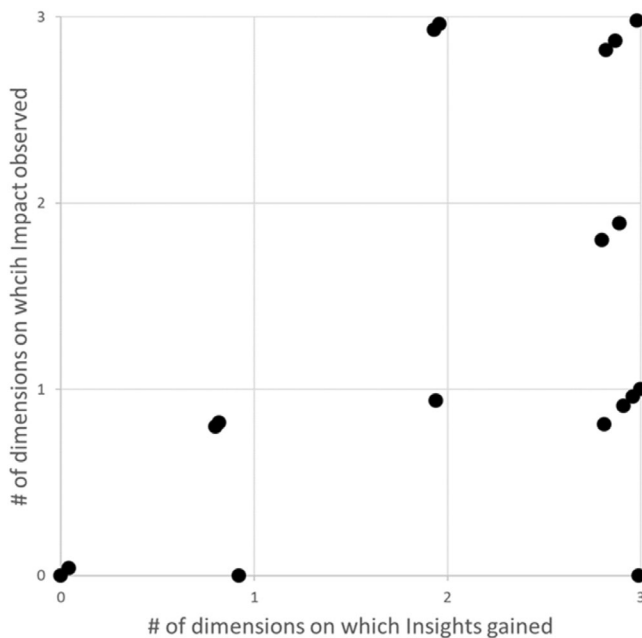
These findings suggests that the share of participants gaining insights is actually a better explanation for the summit impact ( $R^2 = .574$ , see Figure 10) than the share of participants seeing intentions for change ( $R^2 = .347$ , see Figure 6). For example, in case B, the second highest share of participants reported learning (75%), but they were less optimistic about the intentions for change (67%). In the end, they did achieve the highest impact rating (92%).

However, as illustrated by case I, even the link between insights and impact is not deterministic. This



**FIGURE 9** Strongly positive link between share of participants gaining insights and impact on performance realized

summit achieved the highest score of participants gaining new insights (83%), but in the end, only realized limited impact (50%) as the expert observed that implementation challenges made improvement difficult. This finding points again to the role that external factors could play in determining the impact of a summit.



**FIGURE 10** Number of dimensions on which impact realized is only high in cases where insights were gained on multiple dimensions

## DISCUSSION AND CONCLUSION

### Reconsidering the assumed causal process of summits

This study aimed to explicate the assumed causal process of summits and then examine whether this process manifested itself in practice as expected in theory. The longitudinal and systematic evidence from the 18 partnerships analyzed generally lends support to the assumed beneficial effects of summits for collaborative performance. On the whole, a positive impact on performance is realized by summits where a large share of participants learn about the collaboration across a comprehensive set of performance dimensions. In addition, more specific insights are gained about the flow of the process, raising new questions for research.

First, the analysis shows that the links between the various parts of the summit process are not equally tight and direct. There is no strong and traceable link between the performance issues flagged and the insights summit participants gained. Neither the salience of the issue (i.e., the share of participants flagging problems) or the precise content of the issue (i.e., the distribution across the operational, strategic, or constitutional dimensions) seems to be a clear predictor of what participants end up learning. However, the subsequent links between insights and intentions, intentions and impact, or insights and impact do seem to be stronger. This finding suggests that the process linking insights to intentions and impacts is less random.

Second, the analysis shows how the share of participants gaining insights can make or break the success of a summit. This factor best explains the outcome variety of the summits in this dataset (attaining a  $R^2 = .574$ ). And analyzing the pathways of the high versus moderate impact cases suggests that insights gained can be a turning point in the summit process (Figure 8). This finding does not mean that gaining insights is by itself sufficient for achieving impact, that the other steps are unnecessary, or that factors outside the summit process can be discounted. However, this finding does indicate that the quality of the summit as a learning opportunity is highly important.

Finally, the analysis provides a first glimpse into the role that the different operational, strategic, and constitutional dimensions of performance play in the process. There is never a really tight link between the precise performance dimension participants are learning about and the performance dimensions on which they go on to formulate intentions or take action. However, the number of dimensions on which impact is realized is only high in cases where insights were gained on multiple dimensions, suggesting that comprehensive learning is key.

### Reflecting on the limitations of the research design

Before considering the implications of these findings, it is important to acknowledge the limitations of the research design. Ideally, the study would have measured the literacy rates in each municipality before and after the summit, but given the complexity of the issue and the scarcity of data, some proxies for performance impact had to suffice. The support these findings give to the beneficial effects of insightful summits must also be tempered by the fact the 18 cases studied only included municipalities that volunteered to organize a summit. This is an indication of a pre-existing commitment to improvement by the local partners, which probably boosted the chances of achieving impact. Only a study design that compared the impact of summits in willing versus non-willing cases could correct for this selection bias.

The context of this set of cases is also particular. The cases represent collaborations in the literacy field in the Netherlands and the results could potentially be different in other policy domains and countries. Among the group of cases, there could also be more variation than captured in the current research design. The cases were similar with respect to many key factors, including country setting and policy framework, but could still differ in the precise structure, culture, and composition of the collaboration. Perhaps these context variables help to explain the gap between the insights and intentions at the end of the summit and the gap between insights gained and impact realized after the summit.

Finally, the research design makes use of multiple surveys spread out over three points in time and using different respondent groups. Surveying participants before the summit and at the end of the summit, and surveying regional experts months after the summit allows for tracking the summit process over time, but also introduces the risk of outside events shaping the process. Furthermore, asking literacy experts rather than the participants themselves to rate performance impact is a good strategy for avoiding common source bias and self-congratulatory assessments, but this data strategy does mix the assessments of different respondents who may have different yardsticks for measuring performance.

## Exploring questions for future research

In the light of the findings, several questions for future research emerge. A first question concerns the scattershot connection between the issues flagged pre-summit and the insights gained, intentions made, and impact achieved. This loose connection might reflect the innate randomness of social processes surrounding these summits (Moynihan, 2006), the emergent nature of collaborations (Ansell & Gash, 2008), the lack of a fixed script for addressing wicked problems (Head & Alford, 2015), or the dynamic nature of performance management (Bianchi, 2016). From these perspectives, collaborations improve not through linear issue-insight-intervention-impact trajectories, but rather through actors gaining insights and fixing what they can. Perhaps the seemingly scattershot translation between information and conclusions is not a bug, but rather a fundamental feature of effective interactive performance routines?

A second question concerns the importance of learning about operational, strategic, and constitutional dimensions. There was no direct link between dimensions participants were learning about and the dimensions partners took action on, but the number of dimensions on which impact was realized was only high in the cases where participants gained insights on multiple dimensions. This effect could simply be about probability (you are more likely to realize impact on at least one dimension if you aim at multiple dimensions), but could also be a product of the fundamentally interconnected nature of operational service delivery, strategic goals, and constitutional dynamics in collaborations (Emerson & Nabatchi, 2015; Provan & Milward, 2001). Can the operations, strategies, and constitutions of collaborations ever be changed, or studied, separately from each other?

Finally, the data on the different trajectories of the various partnerships asks for a further exploration of the different configuration of conditions that allow for a summit to be impactful. The current analysis of the data suggests that there is a positive relationship between insights gained and impact observed (the more participants report insight, the more impact is achieved).

Although insight seem necessary for achieving impact, the fact that not all summits with high insights go on to realize high impact suggests that gaining insights is not sufficient. More complete explanations could be explored by zooming out from the summit process and considering the role of contextual factors. How do the characteristics of a specific collaborative performance summit combine with the characteristics of the wider collaborative performance regime to make interactive performance routines impactful?

## ACKNOWLEDGMENTS

The authors thank all the people from the National Literacy Foundation, municipalities, and local organizations that participated in the summits. We would also like to thank Rieneke Prins for collecting data about the impact of the summits. Finally, we would like to thank the various colleagues and anonymous reviewers who commented on earlier drafts of this article and helped to make it much better.

## FUNDING INFORMATION

The organization of the summits was financed by the National Foundation for Literacy in the Netherlands (Stichting Lezen en Schrijven). Further research was financed by an Idea Generator grant (NWA.1228.192.190) of the Dutch Research Council (NWO).

## REFERENCES

- Ansell, Chris, and Alison Gash. 2008. "Collaborative Governance in Theory and Practice." *Journal of Public Administration Research and Theory* 18(4): 543–71.
- Behn, Robert D. 2014. *The PerformanceStat Potential: A Leadership Strategy for Producing Results*. New York: Brookings Institution Press.
- Bianchi, C. 2016. *Dynamic Performance Management*, Vol 1. Berlin: Springer.
- Bryson, J. M., and B. C. Crosby. 1993. "Policy Planning and the Design and Use of Forums, Arenas, and Courts." *Environment and Planning B: Planning and Design* 20(2): 175–94.
- Bryson, J. M., B. C. Crosby, and D. Seo. 2020. "Using a Design Approach to Create Collaborative Governance." *Policy & Politics* 48(2020:1): 167–89.
- Bryson, J. M., B. C. Crosby, and M. M. Stone. 2015. "Designing and Implementing Cross-Sector Collaborations: Needed and Challenging." *Public Administration Review* 75(5): 647–63.
- Choi, I., and D. Moynihan. 2019. "How to Foster Collaborative Performance Management? Key Factors in the US Federal Agencies." *Public Management Review* 21(10): 1538–59.
- Collier, D. 2011. "Understanding Process Tracing." *PS: Political Science & Politics* 44(4): 823–30.
- Douglas, S., and C. Ansell. 2021. "Getting a Grip on the Performance of Collaborations: Examining Collaborative Performance Regimes and Collaborative Performance Summits." *Public Administration Review* 81(5): 951–61.
- Douglas, S., C. Ansell, C. F. Parker, E. Sørensen, P. T Hart, and J. Torfing. 2020. "Understanding Collaboration: Introducing the Collaborative Governance Case Databank." *Policy and Society* 39(4): 495–509.
- Douglas, S., and M. J. Schifflers. 2021. "Unpredictable Cocktails or Recurring Recipes? Identifying the Patterns that Shape Collaborative Performance Summits." *Public Management Review* 23(11): 1705–23.
- Emerson, K., and T. Nabatchi. 2015. "Evaluating the Productivity of Collaborative Governance Regimes: A Performance Matrix." *Public Performance and Management Review* 38(4): 717–47.



- George, A. L., and A. Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge: MIT Press.
- George, B., R. M. Walker, and J. Monster. 2019. "Does Strategic Planning Improve Organizational Performance? A Meta-Analysis." *Public Administration Review* 79(6): 810–9.
- Gerlak, Andrea K., and Tanya Heikkila. 2011. "Building a Theory of Learning in Collaboratives: Evidence from the Everglades Restoration Program." *Journal of Public Administration Research and Theory* 21(4): 619–44.
- Gerrish, Ed., ed. 2016. "The Impact of Performance Management on Performance in Public Organizations: A Meta-Analysis." *Public Administration Review* 76(1): 48–66.
- Head, B. W., and J. Alford. 2015. "Wicked Problems: Implications for Public Policy and Management." *Administration & Society* 47(6): 711–39.
- James, O., A. L. Olsen, D. Moynihan, and G. G. Van Ryzin. 2020. *Behavioral Public Performance: How People Make Sense of Government Metrics*. Cambridge: Cambridge University Press.
- Koppenjan, J. F. M., and E. H. Klijn. 2004. *Managing Uncertainties in Networks: A Network Approach to Problem Solving and Decision Making*. London: Routledge.
- Kroll, A. 2002. *Shared Measures: Collective Performance Data Use in Collaborations*. Cambridge: Cambridge University Press.
- Laihonen, H., and S. Mäntylä. 2017. "Principles of Performance Dialogue in Public Administration." *International Journal of Public Sector Management* 30(5): 414–28.
- Lindblom, C. E., and D. K. Cohen. 1979. *Usable Knowledge: Social Science and Social Problem Solving*, Vol 21. New Haven: Yale University Press.
- Moynihan, Donald P. 2005. "Goal-Based Learning and the Future of Performance Management." *Public Administration Review* 65(2): 203–16.
- Moynihan, D. P. 2006. "What Do we Talk about when we Talk about Performance? Dialogue Theory and Performance Budgeting." *Journal of Public Administration Research and Theory* 16(2): 151–68.
- Moynihan, D. P., and A. Kroll. 2016. "Performance Management Routines that Work? An Early Assessment of the GPRA Modernization Act." *Public Administration Review* 76(2): 314–23.
- OECD. 2019. *Survey of Adult Skills 2018*. Paris: OECD.
- Provan, K. G., and H. B. Milward. 2001. "Do Networks Really Work? A Framework for Evaluating Public-Sector Organizational Networks." *Public Administration Review* 61(4): 414–23.
- Schneider, C. Q., and C. Wagemann. 2010. "Standards of Good Practice in Qualitative Comparative Analysis (QCA) and Fuzzy-Sets." *Comparative Sociology* 9(3): 397–418.
- Van Dooren, W., G. Bouckaert, and J. Halligan. 2015. *Performance Management in the Public Sector*. London: Routledge.
- Vangen, Siv, and Chriss Huxham. 2013. "Building and Using the Theory of Collaborative Advantage." In *Network Theory in the Public Sector: Building New Theoretical Frameworks*, edited by Robyn Keast, Myrna P. Mandell, and Robert Agranoff, 51–69. New York: Routledge.
- Weick, K. E. 1995. *Sensemaking in Organizations*. London: Sage.
- Yin, R. K. 2009. "How to Do Better Case Studies." *The SAGE Handbook of Applied Social Research Methods* 2: 254–82.

## AUTHOR BIOGRAPHIES

**Scott Douglas** is an Associate Professor of Public Administration at the Utrecht School of Governance. His research focuses on the performance management of collaborations, working closely with public sector organizations tackling issues such as radicalization, educational inequality, and obesity.

**Christopher Ansell** is a Professor of Political Science at the University of California, Berkeley. His research focuses on understanding how organizations, institutions, and communities can engage effectively in democratic governance in the face of conflict, uncertainty, and complexity.

**How to cite this article:** Douglas, Scott, and Chris Ansell. 2023. "To the Summit and Beyond: Tracing the Process and Impact of Collaborative Performance Summits." *Public Administration Review* 83(5): 1108–1122. <https://doi.org/10.1111/puar.13598>

**APPENDIX A: DETAILED RESULTS PER CASE**

Cases	Issues flagged before the summit			Insights gained at the summit			Intentions for change at end of summit			Impacts observed after summit		
	Operational	Strategic	Constitutional	Operational	Strategic	Constitutional	Operational	Strategic	Constitutional	Operational	Strategic	Constitutional
A	60%	79%	55%	80%	20%	40%	20%	20%	60%	50%	75%	50%
B	68%	60%	15%	75%	75%	75%	75%	75%	50%	100%	75%	100%
C	75%	68%	38%	33%	67%	67%	67%	67%	67%	50%	50%	75%
D	37%	43%	4%	83%	50%	50%	50%	83%	33%	100%	75%	75%
E	31%	44%	28%	63%	63%	63%	63%	38%	75%	50%	75%	50%
F	36%	43%	29%	29%	0%	0%	0%	29%	29%	25%	25%	0%
G	33%	30%	20%	58%	83%	58%	83%	58%	75%	50%	75%	50%
H	55%	67%	40%	50%	100%	75%	100%	100%	50%	75%	75%	50%
I	69%	71%	50%	100%	50%	100%	50%	50%	75%	50%	75%	25%
J	35%	36%	14%	83%	83%	50%	83%	50%	100%	75%	75%	75%
K	62%	68%	27%	56%	67%	56%	67%	89%	56%	50%	50%	50%
L	70%	58%	60%	80%	40%	60%	40%	60%	60%	75%	75%	75%
M	62%	78%	53%	50%	33%	17%	33%	67%	33%	25%	50%	75%
N	30%	33%	8%	67%	83%	50%	83%	83%	50%	50%	75%	0%
O	56%	55%	48%	40%	100%	80%	100%	40%	100%	75%	75%	75%
P	45%	64%	38%	75%	100%	50%	100%	75%	100%	75%	75%	50%
Q	24%	18%	5%	56%	33%	22%	33%	0%	0%	50%	50%	25%
R	35%	27%	16%	44%	22%	22%	22%	44%	33%	25%	25%	25%